Assignment 4: Everyone Has A Number

2020 Fall EECS205002 Linear Algebra

Due: 2021/1/18

Suppose we want to give each student a number based on his/her personal data, such as weight and hight. How can we do that? One simple strategy is to find a weighted sum,

$$s = \alpha_1 w + \alpha_2 h,$$

that combines both factors. The question is how to decide the ratio between α_1 and α_2?

One solution is to perform the Principle Component Analysis (PCA) of the given data. Suppose we have N students, whose weights and heights are (w_1, h_1), (w_2, h_2), $\ldots$, (w_N, h_N). The PCA computes a vector (1-D subspace), so that the data projected to it have the maximum variance, as shown in Figure 1. The solution of the vector is called the principle component of the given data.

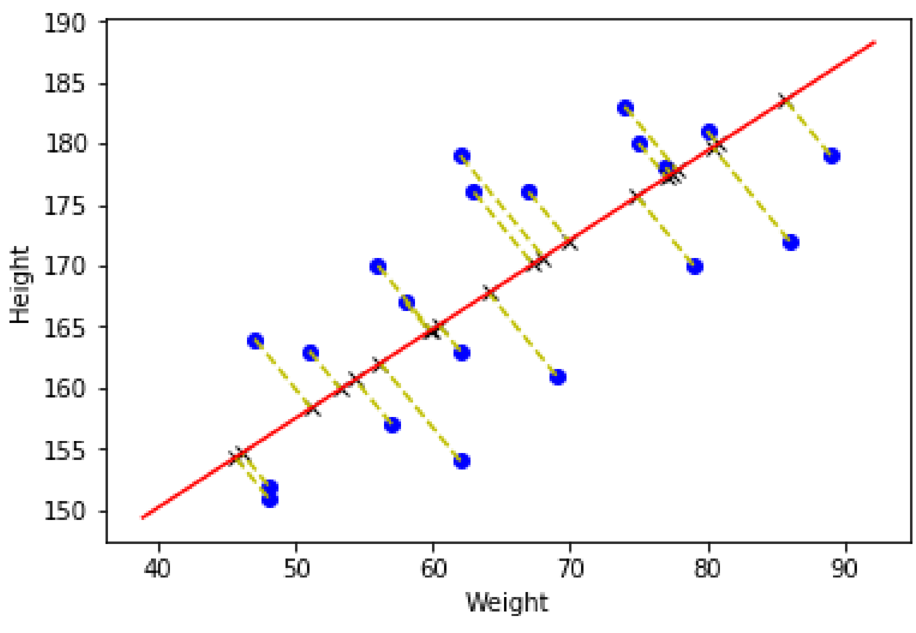

Figure 1: The principle components of the given data.

Now let's translate this problem to the language of linear algebra. Let $x_i = (w_i, h_i)^T$ and the principle component be $y = (\alpha_1^*, \alpha_2^*)$, where $\|y\| = 1$. Recall the projection of x_i to y is (see Section 5.1)

$$Px_i = (yy^T)x_i = (y^T x_i)y = a_i y.$$

where $P = yy^T$ is the projection matrix, and $a_i = y^T x_i$ is the coordinate of x_i's projection onto y. Now the problem of PCA becomes a one-dimensional variance maximization problem. Recall the basic statistics, the mean and the variance of $a_1, a_2, \ldots, a_N$ are

$$\bar{a} = \frac{1}{N}\sum_{i=1}^{N} a_i \text{ and } \sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(a_i - \bar{a})^2.$$

So the problem of PCA can be expressed as

$$\max_{y, \|y\|=1} \frac{1}{N}\sum_{i=1}^{N}(a_i - \bar{a})^2 \tag{1}$$

It can be shown that (1) is equivalent to the following expression (why?)

$$\max_{y, \|y\|=1} \frac{1}{N}\sum_{i=1}^{N} \|P(x_i - \bar{x})\|^2 \tag{2}$$

where $\bar{x}$ is the mean of x_i.

Let's continue to work on the details of (2). The square of vector norm equals to the inner-product of itself, and $P = yy^T$ is the projection matrix. So

$$\begin{aligned}\|P(x_i - \bar{x})\|^2 &= (x_i - \bar{x})^T P^T P (x_i - \bar{x}) \\ &= (x_i - \bar{x})^T yy^T yy^T (x_i - \bar{x}) \\ &= (x_i - \bar{x})^T yy^T (x_i - \bar{x})\end{aligned}$$

We can do that because $y^T y = \|y\|^2 = 1$.

Next, since $(x_i - \bar{x})^T y$ and $y^T (x_i - \bar{x})$ are scalars, we can exchange those two terms in their product and rewrite the above equation as

$$y^T (x_i - \bar{x})(x_i - \bar{x})^T y.$$

Let's define the *covariance matrix* for the vector data x_i as

$$\Sigma = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(x_i - \bar{x})^T. \tag{3}$$

For $x_i = (w_i, h_i)^T$, the covariance matrix is

$$\Sigma = \frac{1}{N}\sum_{i=1}^{N}\begin{bmatrix}(w_i - \bar{w})^2 & (w_i - \bar{w})(h_i - \bar{h}) \\ (w_i - \bar{w})(h_i - \bar{h}) & (h_i - \bar{h})^2\end{bmatrix}$$

where $\bar{w}$ and $\bar{h}$ are the means of w_i and h_i respectively. With Σ, equation (2) can be expressed as

$$\max_{y, \|y\|=1} y^T \Sigma y. \tag{4}$$

How to solve the maximization problem (4)? Let's use the knowledge in Section 6.4 to help. First, you can observe that Σ is symmetric, which means $\Sigma = \Sigma^T$. So we can find an orthogonal matrix U that diagonalizes Σ,

$$\Sigma = U \Lambda U^{-1} = U \Lambda U^T,$$

where Λ is a diagonal matrix with diagonal elements $\lambda_1, \lambda_2, \ldots, \lambda_N$. With the above eigen-decomposition, $y^T \Sigma y$ can be written as

$$y^T \Sigma y = y^T U \Lambda U^T y = z^T \Lambda z.$$

Since $\|y\| = 1$ and U is orthogonal, $\|z\| = \|U^T y\| = 1$, (see Section 5.5).

For our 2D problem, let $z = (z_1, z_2)^T$ and $\Lambda = \text{diag}(\lambda_1, \lambda_2)$. So the problem of (4) is equivalent to

$$S = \max_{z_1, z_2, z_1^2 + z_2^2 = 1} \lambda_1 z_1^2 + \lambda_2 z_2^2 \tag{5}$$

From the equivalence of (1) and (4), you can show that $\lambda_i \geq 0$ for $i = 1, 2, \ldots, N$. (why?) Suppose $\lambda_1 > \lambda_2 \geq 0$, the solution of (5) is

$$z_1 = 1, z_2 = 0, \text{ and } S = \lambda_1. \text{ (why?)}$$

From the definition $z = U^T y$,

$$z = \begin{bmatrix} 1 \\ 0 \end{bmatrix} = U^T y = \begin{bmatrix} u_1^T y \\ u_2^T y \end{bmatrix}.$$

This implies y is the first column vector of $U = [u_1 \ u_2]$, because $u_1^T y = 1$ and $u_2^T y = 0$.

We can extend this result to any finite dimensions. The solution of PCA is the eigenvector u_1 of the covariance matrix Σ, corresponding to the largest eigenvalue λ_1, $\Sigma u_1 = \lambda_1 u_1$. The maximum variance of the projected value is λ_1.

1 Assignments

1. (20%) Show that (1) and (2) are equivalent. (Hint: to show that $\bar{a} = P\bar{x}$.)

2. (20%) Show all the eigenvalues of the covariance matrix Σ are nonnegative. (Hint: from the equivalence of (2) and (5) to show that $y^T \Sigma y \geq 0$ for any y, and then show this fact implies all the eigenvalues of Σ are nonnegative.)

3. (20%) Let the eigenvalues of Σ be $\lambda_1 > \lambda_2 \geq \cdots \geq \lambda_N \geq 0$. Show that the solution of (4) must be $y = u_1$, the eigenvector corresponding to λ_1.

4. (20%) For each student, we have collect his/her personal data $x_i = (w_i, h_i, z_i)$. `https://forms.gle/3LZ1LKrAaJCZ3C1z6`

   (a) (10%) Remove your own data from the list, and use other data to compute the PCA. Then use the mean and the principle component of other data to calculate your number.

   (b) (10%) Remove your own data from the list, and use other data to compute the PCA of 2D data first, (w_i, h_i), and find the weighted sum $t_i = \alpha_1 w_i + \alpha_2 h_i$. Then compute the PCA of (t_i, z_i). Plot the figure of t_i, z_i, and their principle components, as Figure 1. Compare the results with (b). Will they give the same numbers? Discuss the reasons. (If they are the same, why? And if they are different, why?)

5. (20%) How to project the 3D data onto a 2D subspace so the variance of projected data are maximized? Design an algorithm, prove its correctness, and write a python code using the given data to show your algorithm works pictorially. Your algorithm needs to find an orthogonal basis of the 2D subspace $Y = [y_1 \ y_2]$, $Y^T Y = I_2$. The definition of 2D variance is

$$\bar{a} = \frac{1}{N} \sum_{i=1}^{N} a_i, \ \sigma^2 = \frac{1}{N} \sum_{i=1}^{N} \|a_i - \bar{a}\|^2, \tag{6}$$

   where a_i is the projected x_i from 3D to the 2D subspace, and $\bar{a}$ is their center (mean). The problem is to maximize σ^2.

2 Submissions

1. Write a report in a PDF file that includes (1), (2), (3), (4), and (5). For question (4), attach the plots and your discussion. For question (5), give your algorithm and proofs.

2. Python code of your implementation of (4) and (5).

3. Zip them and submit to iLMS system.